Padmakala S* et al.
(IJITR) INTERNATIONAL JOURNAL OF INNOVATIVE TECHNOLOGY AND RESEARCH
Volume No.7, Issue No.5, August-September 2019, 9317-9332.

A Literature Study On Video Retrieval Approaches

PADMAKALA S
Department of Computer Science & Engineering, St. Joseph's Institute of Technology, Chennai

Abstract: **A detailed survey has been carried out to identify the various research articles available in the literature in all the categories of video retrieval and to do the analysis of the major contributions and their advantages, following are the literature used for the assessment of the state-of-art work on video retrieval. Here, a large number of papershave been studied.**

Key words: **Video Retrieval; Analysis; Assessment; Content-Based**

I. INTRODUCTION

A detailed survey has been carried out to identify the various research articles available in the literature in all the categories of video retrieval, and to do the analysis of the major contributions and their advantages. Following are the literatures used for the assessment of the state-of-art work on video retrieval. Here, sixty-five works have been analyzed.

1. Content based Video Retrieval System

In this section, we explain the some of the paper related to content-based video retrieval system. Simon Jones & Ling Shao (2013) explained the Content-based retrieval of human actions from realistic video databases. Here, several recently developed action representation and information retrieval techniques in a human action retrieval system were presented. These techniques include various means of local feature extraction; soft-assignment clustering; Bag-of-Words, vocabulary guided and spatial-temporal pyramid matches for action representation; SVMs and ABRS-SVMs for relevance feedback. Successful application of relevance feedback, in particular, resulted in more practical systems.Moreover, Ja-Hwung et al. (2010) have explained the Effective content-based video retrieval using pattern-indexing and matching techniques. Recently, multimedia data has grown rapidly due to the advanced multimedia capturing devices, such as digital video recorder, mobile camera and so on. In addition, video retrieval using query-by-image was also not successful in associating the videos with user’s interest. An innovative method was employed to achieve the high quality of content-based video retrieval to discover temporal patterns in video contents. On basis of the discovered temporal patterns, an efficient indexing technique and an effective sequence matching technique were integrated. This reduced the computation cost to raise the retrieval accuracy.

Similarly, Anastasios et al. (2000) have introduced a fuzzy representation of visual content. This was useful for the new emerging multimedia applications, such as content-based image indexing and retrieval, video browsing and summarization. A multi-dimensional fuzzy histogram was constructed for each video frame based on a collection of appropriate features, extracted using video sequence analysis techniques. This approach was applied for both video summarization and content based video indexing and retrieval. In the initial case, video summarization was accomplished by discarding shots or frames of similar and meaningful visual content is retained (key-frames). In the second phase, a content-based retrieval scheme is investigated, so that the most similar images to a query are extracted. Moreover, Sebastien et al. (2003) explained the real-time segmentation of uncompressed video sequences for content-based search and retrieval.Dyana & Das (2010) have explained the Spatial-Temporal Representation for Content-Based Video Retrieval. Features from spatial, along with temporal information were integrated into a unified framework.This helped to retrieve similar video shots. A sequence of orthogonal processing was done using a pair of 1-D multi-scale and multispectral filters, on the space-time volume (STV) of a video object (VOB) is responsible for producing a gradually evolving (smoother) surface. Zero-crossing contours (2-D) were computed using the mean curvature. This evolving surface was stacked in layers to yield a hilly (3-D) surface, for a joint multi spectral-temporal curvature scale space (MST-CSS) representation of the video object. Peaks and valleys (saddle points) were detected on the MST-CSS surface which was further used for feature representation and matching. Computation of the cost function for matching a query video shot with a model involves matching a pair of 3-D point sets.

Likewise, Haojin Yang & Meinel (2014) explained the automated video indexing and video search in large lecture video archives. Firstly, automatic video segmentation and key frame detection are applied to offer a visual guideline for the video content navigation. Subsequently, textual metadata is extracted by applying Automatic Speech Recognition (ASR) on lecture audio tracks and video Optical Character Recognition (OCR) technology on key-frames. The OCR and ASR transcript, as well as detected slide text line types, were adopted for keyword extraction. These were the means by which both video- and segment-level keywords were extracted for content-based video browsing and search.

2. Text based Video Retrieval

Extraction of text from image and video was an important step which helps to build efficient indexing and retrieval systems for multimedia databases. Sudhir & Pradeep (2012) explained the Lawn Tennis Video Summarization based on Audiovisual and Text Feature Analysis. The method uses frame color histogram to classify video into two namely-play field color shots (PFCS) and non-play field color shots (NPFCS). Playfield color shots were the segments of interest used to recognize the tournament class. A dominant colored frame from every PFCS was a salient feature in this frame. Their approach also employs dominant values of PFCS volume and energy. For each, PFCS dominant audio energy value was computed and then the corresponding key frame was extracted. On-screen text which is presented in playfield color shots provides important semantic information so the last frame of the shot is extracted as a key caption. Moreover, Lei Xu & Kongqiao (2008) explained the Extracting text Information for Content-based Video Retrieval. In the detection stage, edge density feature, pyramid strategy, and some weak rules were utilized to search for text regions, so that high detection rate was achieved. Meanwhile, to eliminate the false alarms and to improve the precision rate, a multilevel verification strategy was adopted. In the segmentation stage, a precise polarity estimation algorithm was provided at first. Then, multiple frames containing the same text is integrated to enhance the contrast between text and background.In Chinmaya Misra et al. (2006) have explained the Content-Based Image and Video Retrieval Using Embedded Text. A hybrid approach was adopted for text extraction by exploiting a number of characteristics of text blocks in color images and frames of the videos. This system helps to detect both caption text as well as scene text of different color, intensity, font and size. They developed an application for on-line extraction and recognition of texts from videos which were used for video clips retrieval based on keyword.

3. Motion based Video Retrieval

Here, we consider some of the work related to video retrieval based on motion. In Jun Wei Hsieh et al. (2006) have explained a hybrid motion based system to retrieve the preferred videos from video databases by means of trajectory matching. Here, sketch-based and string-based schemes are utilized to analyze and index the trajectory with more syntactic meanings. Initially, a set of control points is extracted from each trajectory as features using a sampling method. Then, some missed data in the set of control points have been interpolated using a curve fitting approach in the sketch-based method. Subsequently, the visual distance between any two trajectories is measured directly by comparing their position data. Also, the string-based scheme has been employed in addition to the visual distance to compare any two trajectories based on their syntactic meanings. Moving object tracking technology was utilized for content based video retrieval, Che-Yen Wen et al. (2007). Here, the moving pixels were detected using background subtraction, which has defeated the shadow problem. The noise was barred and the moving pixel was ameliorated by means of connected components labeling and morphological operations. The target image and information for content base video retrieval in the database have is extracted with the help of color histograms, color similarity, and “motion vector”. With the proposed technique, the image frame retrieval is performed in multi-CCD surveillance systems or single-Charge Coupled Device (CCD). During multi-surveillance detection, retrieval, and retrieval error occurred due to sudden environment changes

Moreover, Dagtas et al. (2000) explained the motion-based video indexing and retrieval.Another reason for the use of such techniques was the need for efficient query processing in real-time applications. They presented models that use the object motion information. This was done in order to characterize the events to allow subsequent retrieval. An algorithm for different spatiotemporal search cases was developed.This was done using various signal and image processing techniques. They developed a prototype video search engine called PICTURESQUE to verify the above methods.Additionally, Su et al. (2007) explained the Motion Flow-Based Video Retrieval. The use of motion vectors embedded in MPEG bit streams was used to generate so-called ldquo motion flow srdquo. This is applied to perform video retrieval. By direct usage of the motion vectors, there does not arise the need to consider the shape of a moving object and its corresponding trajectory. Instead, the simple ldquolinkrdquo, the local motion vectors across consecutive video frames which formed motion flows, were then recorded and stored in a video database. In the video retrieval phase, a new matching strategy was proposed to execute the video retrieval task. Motions that do not belong to the mainstream motion flows are filtered out by their algorithm.

4. Audio based Video Retrieval

Basically, the video has the audio and texture information. Based on the audio feature we extract the video. In Babaguchi et al. (2004) has explained the Scene retrieval with the matching of sign sequence using video and audio features. This paper presents a method of similar scenes retrieval from the video with the technique of sign sequence matching. The visual and auditory streams were divided into fixed-length video/audio packets. Using the features of packets extracted, sign sequences were formed which was viewed as an abstraction of video and audio features. DP which matches between the target and query sign sequences allows us to find scenes similar to the query in the video stream. In Liu Huayong (2004) has explained the content-based video retrieval, a kind of retrieval by its semantical contents. Since video data is composed of multimodal information streams such as visual, auditory and textual streams, a strategy for automatic parsing sports video using multimodal analysis is described. The paper explains the structure of sports video database system and then goes to introduce a new approach that integrates visual streams analysis and speech signal processing and text extraction to realize video retrieval. Event detection using audio-visual features in field sports video using audio-visual features and a support vector Machine was explained in Sadlier et al. (2005).

Features that indicate significant events were selected and robust detectors were built. These features were rooted in characteristics that are common to all genres of field sports. The evidence gathered by the feature detectors was combined by means of a support vector machine. The system was tested generically across multiple genres of field sports including hockey, soccer, rugby and Gaelic football. Similarly, Yipei Wang et al. (2014) explained the Exploring audio semantic concepts for event-based video retrieval. This paper presented a novel framework to handle the complex situation of semantic information extraction in real-world videos.It is evaluated through the NIST multimedia event detection task (MED). The occurrence confidence matrix of sound events was calculated and multiple strategies were explored to generate clip-level semantic features from the matrix. For detecting speech segments Short-Time Energy and MFCC is used. Pitch and pause rate are responsible for detection of excited speech.

5. Events based Video Retrieval

Event retrieval was one of the important topics of research in multimedia processing. Several works related to the retrieval of video based on events were presented here; In Ganesh & Dipali (2014) have explained the Review on Event Retrieval in Soccer Video. Their method aims at the efficient building of human knowledge directly for soccer video events retrieval by fuzzy systems. Moreover, Changsheng Xu (2008) explained the sports video semantic event detection based on analysis and followed by alignment of Webcast text. Webcast text was used as a text broadcast channel for sports game which was co-produced with the broadcast video. This is easily obtained from the Web. First, the Webcast text was analyzed to cluster and detect events based on text in a simple way using probabilistic latent semantic analysis (PLSA). Based on the detected text event and using video structure analysis, a conditional random field model (CRFM) is employed to align text event and video event.This was done by detecting the event moment and event boundary in the video.

Similarly, Vijayakumar & Nedunchezhian (2012) have explained the Event detection in cricket video based on visual and acoustic features. In this paper, video and audio features were presented based on event detection approach. This was shown to be effective when applied to the cricket sports video. The advantage of this approach was the ability to recognize events that indicatehigh level of audio response and crowding of players which can be correlated to key events. The event detection has two steps. In the first step, audio and visual features are extracted. Next, by defining a set of heuristic rules, important semantic events like wicket fall, score events etc are detected. In, Deepali Bhawarthi & Shrini (2012) was Garage gave progressed ideas for cricket highlight generation. This was an attempt towards summarization. The key frame detection based approach proved as an outstanding detection correctness. It also resulted in saving the processing time. The categorization reveals an improved exposure and categorization ratio at different stages.

6. Ontology based Retrieval

Kimiaki Shirahama et al. (2007) explained the Content-Based Video Retrieval Using Video Ontology. In this paper; various kinds of events (e.g. conversation, battle, run/walk and so on) are efficiently retrieved from a video archive. At this juncture, a "video ontology" which was a formal and explicit specification of events was constructed, wherein an event was modeled to have 4 dimensions of semantic contents (i. e. Action, Location, Time and Shooting technique). In order to retrieve such events, concepts in 4 dimensions needs to be detected automatically. So, "video data mining" is conducted to extract "semantic patterns" from videos. Here, a semantic pattern, a combination of low-level features (e.g. color; motion and audio) associated with events of a certain kind were used to characterize concepts in 4 dimensions of semantic contents. Furthermore, the video ontology was refined from subspaces of videos by extracting new semantic patterns.Moreover, Nagarajan & Minu (2015) explained the Fuzzy Ontology-Based Multi-Modal Semantic Information Retrieval where the main focus of the paper was improving information retrieval for sports events using Ontologies. Being a complex problem, it was addressed into the following sub-problems in this paper. (1) Integrate domain knowledge and images using fuzzy ontology and to retrieve the required Muti-modal information using fuzzy rule set. (2) Provide image semantic by constructing visual codebook for affine covariant-Semantic segmented patches. (3) Analyze the discriminative power of each visual word using the probabilistic latent semantic and quantizing them using the Chi-Square test.

Similarly, Liya Thomas & Syama (2014) explained the Video Annotation based on Ontology and Retrieval System. The video annotation system was based on ontology with HoG features. They were given for training the classifiers. The SIFT and HoG features of the images were extracted. They were further used for training the classifiers for comparing the classifiers. An analysis of the results helped to find the better feature which helped to train the classifier to get more prominent annotated video database. From the annotated video database, retrieval of the videos based on objects was also done.Alberto Del Bimbo & Marco Bertini (2007) explained the Multimedia Ontology-Based Computational Framework for Video Annotation and Retrieval. The extended ontologies were supposed to support the definition of visual concepts as representatives of specific patterns of a linguistic concept. The linguistic part of the ontology embeds permanent and objective items of the domain and the perceptual part includes visual concepts that are dependent on temporal experience. They are subject to changes with time and perception. This was the reason behind dynamic update of visual concepts to be supported by multimedia ontologies. This helped to represent temporal evolution of concepts.

7. Denoising and Shot Boundary Detection

Xuemei Wang et al. (2015) have explained the structure-adaptive image denoising method with 3D collaborative filtering by optimizing the block matching procedure.This method outperforms the original BM3D in both PSNR and visual quality at the cost of extra computational overhead.Yu Meng et al. (2009) explained the shot boundary detection algorithm based on Particle Swarm Optimization Classifier. This method first takes the difference curves of U-component histograms as a characteristic which takes into account the video frames differences and then utilizes a slide-window mean filter to filter difference curves. A KNN classifier applied PSO to detect and classify the shot transitions. This method has three advantages. It was more transitions sensitivity; each curve graphic with remarkable characteristics corresponds to a shot transition. Cuts and Gradual transitions could be detected in the same step.Moreover, Sun et al. (2011) explained the Novel Shot Boundary Detection (SBD) Method Based on Genetic Algorithm-Support Vector Machine. Initially, features of pixel domain and compressed domain are synthetically extracted. They are then organized into a multi-dimension vector by using the method of sliding window.

To follow, the genetic algorithm was utilized to implement the simulation and iterative optimization. Then the model trained by the approximately optimal parameters was applied to judge and classify the frames of video. Thus SBD was completed to the full. Similarly, Abdelati Malek Amel et al. (2010) have introduced the Video shot boundary detection using motion activity descriptor. The interest in the validation of this descriptor is the aim of its real timeimplementation. It had reasonably high performances in detecting the shot boundary. It used adaptive rood pattern search (ARPS) algorithm to extract the motion activity information in a uncompressed domain. In this context, the motion activity descriptor was applied for different video sequences.In Xiaoming Liu & Tsuhan Chen (2004) have explained the shot boundary detection using temporal statistics modeling. In multimedia information retrieval, the interesting research topic was shot boundary detection. In order to perform shot boundary detection, an algorithm was proposed for modeling temporal statistics using a novel Eigenspace updating method. The feature extracted from the current frame was compared to a model trained from features obtained from the previous frames. A shot boundary was detected if the new feature does not fit well to the existing model. The model was based on the principal component analysis (PCA), or the Eigenspace method. Here the Eigenspace was updated to capture the non-stationary statistics of the features.Moreover Choi et al. (2006) developed an algorithm for automatically segmenting videos into basic shot units. A basic shot unit was taken as an unbroken sequence of frames taken from one camera. At first, they calculate the frame difference by using the local histogram comparison. The dynamical scale of the frame difference was calculated by Log-formula. This was done to compress and enhance the frame difference.

Finally, they detected the shot boundaries by the newly explained shot boundary detection algorithm which was more robust to camera or object motion, and many flashlight events. Likewise Lu & Shi (2013) explained the Fast Video Shot Boundary Detection Based on SVD and Pattern Matching. In the initial stage, the adaptive thresholds were used to predict the positions of the shot boundaries and lengths of gradual transitions. Most non-boundary frames were discarded at the same time. Only the candidate segments that contain the shot boundaries were preserved for further detection. For all frames in each candidate segment, their color histograms in the HSV space was extracted, forming a frame feature matrix. The SVD was then performed on the frame feature matrices of all candidate segments. This helped to reduce the feature dimension. Finally, gradual cut transitions are identified using their pattern matching method based on a new similarity measurement. Out of all the above approaches discussed pixel-wise comparison is easy to be done, it is too sensitive against noise, illumination changes and camera motion.

8. Key Frames based Video Retrieval

Potnurwar & Mohammad (2014) have explained the Video Annotations using Visual Attention Key Frame Extraction. The labeled training data was insufficient to represent the distribution of the entire dataset. This was a major obstacle in the automatic semantic annotation of a large-scale video database, where the objective is to represent the most “important” or “meaningful” scenes of a large amount of visual information. This is to be done by only a few images: the key frames. At First, the image sequences were segmented into segments called shots. Then a few frames of each shot are selected as key frames. Here, retrieving videos using keywords requires the availability of semantic features of the videos. The focus is on visual attention Key frame extraction (VAKE).

Moreover, Zhonghua Sun et al. (2008) have explained the Video Key Frame Extraction Based on Spatial-Temporal Color Distribution. One of the key problems in video content indexing and retrieval was Video key frame extraction. It was a type of video abstraction. At first, a temporally maximum occurrence frame was constructed which considered the spatial and temporal distribution of the pixels throughout the video shot. Then a weighted distance was computed between frames in the shot and the constructed reference frame and key frames were extracted at the peaks of the distance curve. This helped them to achieve high compression ratio and high fidelity.Additionally, Guozhu Liu et al. (2009) have explained the Key Frame Extraction from MPEG Video Stream. Initially, an improved histogram matching method was used for video segmentation. Secondly, the key frames were extracted utilizing the features of I-frame, P-frame, and B-frame for each sub-lens. In order to measure the validity of the method, fidelity and compression ratio was used. Experimental results show that the extracted key frames can summarize the salient content of the video. The method is feasible, highly efficient, and robust.

Similarly, Song & Fan (2006), have explained a joint key frame extraction and object segmentation method. This was done by constructing a unified feature space for both the process, where key frame extraction is done as a feature selection process for object segmentation. This is in the context of GMM-based video modeling.Moreover, Khin Thandar et al. (2013) have explained the Key Frame Extraction for Video Summarization Using Discrete wavelet Transform (DWT) Wavelet Statistics. In extracting key frames, two consecutive frames were first transformed using DWT and then the differences of the detailed components are estimated. If different values of a consecutive pair are higher than the threshold, the last frame of the pair is a key frame. Experimental results also discussed the representation of the validity of the proposed method for video summarization.

In, Lakshmi Priya & Domnic (2014) have explained the Shot based key frame extraction for ecological video indexing and retrieval. Initially, the frames were sequentially clustered into shots. Then the shot frame clustering technique were used. The cluster having a larger dispersion rate was selected for inter-cluster similarity analysis (ICSA) and the sub-shot based keyframes were extracted using ICSA. Likewise, Kin-Wai Sze et al. (2005) have explained optimal key frame representation scheme based on global statistics for video shot retrieval. Each pixel in this optimal keyframe was constructed after taking into consideration the probability of occurrence of those pixels at the corresponding pixel position. This was among the frames in a video shot. Therefore, this constructed key frame was called temporally maximum occurrence frame (TMOF), which was an optimal representation of the frames in a video shot. The performance of this representation scheme was further improved by considering the k pixel values. Those with the largest probabilities of occurrence and the highest peaks of the probability distribution of occurrence at each pixel position for a video shot was considered. The corresponding schemes ware called k-TMOF and k-pTMOF, respectively. The key frame constructed using the above procedure gave an optimal solution. But the changes in illumination of the scene gave less retrieval rate.

Moreover, Wattanarachothai & Patanukhom (2015) explained the frame extraction for text based video retrieval by means of Maximally Stable Extremal Regions. The scheme comprises of three main processes that were key frame extraction and text localization, followed by keyword matching. For the key-frame extraction, a Maximally Stable Extremal Region (MSER) based feature was explained which was oriented to segment shots of the video with different text contents. In this process, to form the text lines, the MSERs in each key frame are clustered. This is based on their similarity in stroke width, position, size, and then color. Subsequently, Tesseract OCR engine was used for recognizing the text regions.

9. Video and Audio Feature Extraction

Drelie Gelasca & Ebrahimi (2009) have explained a perceptually determined target metric for segmentation quality assessment, in light of psychophysical tests on manufactured artifacts. The perceptual metric proposed was tried by directing a study on the real artifacts delivered by a run of the mill video object segmentation. Seven late segmentation calculations were picked and dissected both equitably and subjectively. At first, the real artifacts presented were characterized in view of a subjective perception. Furthermore, they proposed a perceptual target metric equipped for foreseeing the subjective quality as saw by human viewers. In order to generate a ground truth for the evaluation of motion-based algorithms for video object segmentation a new design procedure was employed by Tiburzi et al. (2008). On a general basis, motion-based segmentation algorithms are either based on a few pixel intensity models particularly on some optical flow estimation of the scene when dealing with moving cameras.

A Segmentation method for deformable objects in monocular videos was presented by Peng Tang & Lin Gao (2008). Initially, they introduced a dynamic shape to symbolize the prior knowledge about object shape deformation in a manner of the auto-regressive model. Here the shape is treated as a function of subspace shapes at previous time steps. A framework of Markov random field energy was developed by fusing both spatial-temporal image information and model predictions. The devised framework of Markov random field energy can be effectively minimized by the graph cut algorithm. Thus a global optimum segmentation is achieved. Both the orthogonal basis and the autoregressive model parameters are updated on-line, using final segmentation results in order to capture model variations. This helps in forming an effective closed loop system. Lastly, the use of the proposed segmentation method with respect to noise, clutter, and partial occlusions were demonstrated by means of promising experimental results.

In Dubravko Culibrk et al. (2007) presented a background modeling and subtraction approach for video object segmentation. In correspondence to their application domain, they also proposed Neural Network (NN) architecture with the intention of building an unsupervised Bayesian classifier. The segmentation in natural-scene sequences with complex background motion and changes in illumination were efficiently handled by constructed classifier. The weight of the proposed NN acts as a model of the background and is updated temporally to reflect the observed statistics of background. The test pool contains diverse surveillance-related sequences, which was formerly published. Based on this, the segmentation performance of the proposed NN was examined both qualitatively and quantitatively.Moreover, Aree A. Mohammed et al. (2008) has developed an object extraction method and proposed efficient algorithms to characterize object motion. The set of their proposed tools serves as the base for development of object based functionalities. This was used for manipulating video content. The estimators produced different algorithms where the comparison was done with respect to quality and performance and was also tested on real video sequences. Their proposed method is an aid for the most recent encoding standards and description of multimedia content – MPEG4 and MPEG7.

Ming Zhao et al. (2002) elaborated on the Semi-Automatic Video Object Segmentation Basing on Hierarchy Optical Flow. The segmentation method consists of spatial and temporal segmentation. For the spatial segmentation, a point-based graphic user interface (PBGUI) was presented. Here the user could input easily, active contour model and tracking. Bug algorithm was applied to define precisely the segmentation of the video object of interest. As a result of spatial segmentation, the temporal segmentation involves non-rigid object contour tracking and rigid object whole-tracking. This was done by hierarchy optical flow algorithm based on the Lucas-Kanade algorithm. And the tracking point selection algorithm was explained at length to improve the tracking performance in the rigid object whole-tracking.Similarly, Jordi Pont et al. (2015) explained the Semi-Automatic Video Object Segmentation by Advanced Manipulation of Segmentation Hierarchies. This paper presented an interface based on a click-and-drag interaction that allows rapid selection of regions from state-of-the-art segmentation hierarchies. The interface had a good response and helped to obtain very accurate segmentation. It was designed to minimize the human interaction. To evaluate the results, they provided a new set of object video ground truth data.

Additionally, Chasanis et al. (2007) explained the Scene detection in videos using shot clustering and symbolic sequence segmentation. Here, the video was first segmented into shots and key frames. Then they were extracted using the global k-means clustering algorithm which represented each shot. Further, an improved spectral clustering method was applied to cluster the shots into groups based on visual similarity. A label was assigned to each shot according to the group it belonged to. Next, a method for segmenting the sequence of shot labels is applied. The final scene was provided as a segmentation result. Moreover, Sankar et al. (2006) solved the difficulties in the temporal segmentation of videos. To carry out the segmentation, they offered a multi-modal approach where hints from different data sources were combined. Here, the video was segmented to meaningful articles or sight, by the scene level explanations presented by the commentary. The segments were then automatically interpreted with the relevant explanations. This allows for a semantic contract and recovery of video segments. This was hard to be attained from available visual feature based approaches.

In numerous applications, for example, picture recovery, video recovery, and video ordering are broadly utilized. A color histogram is the most, for the most part, utilized technique inferable from its power to scaling, introduction, the point of view, and impediment of pictures (Marchiori 1997). The color space models (Chinmaya Misra & Shamik Sural 2006) can be separated as equipment arranged and client situated. The color spaces including RGB and depend on the three-color jolts hypothesis. The RGB color space is a unit block with red, green, and blue tomahawks; henceforth, in RGB color space, a color is spoken to by a vector with three directions. On the off chance that each of the three qualities is set to 0, the relating color is dark. On the off chance that every one of the three qualities is set to 1, the relating color is white (Bole et al. 1998).

The hue, lightness, and saturation (HLS), Hue, Chroma, Value (HCV), HSV, and hue, saturation, and Brightness (HSB) are client arranged color spaces i.e., tint, immersion, and splendour. By and large, RGB, HSV, and luma component, blue difference and red-difference chroma components (Y,Cb,Cr) is the different color models utilized as a part of the color picture. Here, we have utilized Y,Cb,Cr color space that is gotten from the RGB color space. These models indicate the shading substance through 3D subspace, where every subspace contains particular color data. It is conceivable to deliver more color spaces by isolating chromatic data and luminance data.Color minute is one of the color highlights which are generally utilized by the inquiries about. Color minutes are measures that can be utilized to recognize pictures in view of their color. Once registered, these minutes give a measure to the color likeness between pictures. This color likeness can turn into a measure of discovering contrast between back to back edges in a video. The shading circulation of a picture through its minutes is named as a color minute. Color conveyance of an image can be expressively characterized with low request minutes (mean), second request minutes (standard deviation) and third request minutes (unsymmetrical cubic base of the image).

A principle work in the substance depends video look framework is properties extraction. A credit is utilized to catch a specific visual property of a picture. The color is a basic trait for picture exhibit which is widely utilized in picture recovery. This is inferable from the verify that color is invariance regarding picture scaling, interpretation and revolution (Brin & Page 1998). The human eye is vulnerable to hues, and colors properties are one of the mostly basic components encourage people to perceive pictures (Bharat & Henzinger 1998). Color properties are, subsequently, essential attributes of the substance of pictures. Color characteristics extraction strategies by and large fall in two sort viz. worldwide methods and nearby procedures. In worldwide systems, characteristics extraction prepare to think about the whole picture, including worldwide shading histogram, histogram crossing point, picture bitmap, and so forth.

Then again nearby systems think about a segment of the picture, including neighborhood color histogram, color distinction histogram, color correlogram, and so forth. Color Histogram is the most broadly utilized strategy for extricating the color traits of a picture. It compares to the picture from an assorted point of view. It means the dissemination of color containers in a picture. It implies check of undifferentiated from pixels. A color histogram is foreseen as a worldwide color descriptor which examines each measurable color recurrence in a picture. It is utilized to disentangle the issues like change in interpretation, turn, and point of view. Neighborhood color histogram spotlights on the identity parts of a picture. Nearby color histograms esteem the spatial dispersion of pixel which is lost in worldwide color histograms. A color histogram is sans inconvenience to compute and pitiless to little varieties in the picture so is amazingly basic in ordering and recovery of picture database. Steady with the solid relationship amongst hues and human feelings, an enthusiastic semantic inquiry demonstrates rely on upon color semantic portrayal is edified in this fragment.

At first, every edge is restored into HSV color space which is a visual property and practicality in substance depend on picture recovery applications. all things considered, HSV color space encapsulates a visual view of the variety in Hue, Saturation and Intensity estimations of a picture pixel. Thus, HSV qualities are standardized to the range [0, 1]. Tint esteem is much of the time-quantized into a little arrangement of around 10-20 base shading names. We homogeneously quantize the Hue esteem into 10 base hues, water, sea green/blue, yellow, green, red, orange, blue, violet, purple, and red. Immersion and Value are quantized (not consistently) into 4 containers correspondingly as descriptive words symptomatic of the immersion and luminance of the shading. Texture feature is the next video feature to be discussed. A spatial change in the pixel forces or dark qualities is named as the surface. Various specialists have been in the field of study for separating the surface and shading from VSR and an assortment of a few applications are utilized for this exploration. Gordon & Pathak (1999) have presented a powerful surface portrayal technique called the first Local Binary Pattern (LBP) administrator (Salton 1989). For surface depiction, the histogram of the twofold examples figured over an area is utilized. The administrator contracts every pixel and the dark levels of its neighbouring pixels.

10. Process of LBP operator

In this strategy, each image is divided into an arrangement of squares. The edge capacity is utilized by the LBP administrator to the arrangement of squares for indicating the pixel values. This edge capacity is performed on the 3×3 neighborhood of every last pixel by a method for the inside esteem. Every last pixel which is situated in the piece has eight neighbors, for example, left-best, left-center, left-base, right-best, and so on. These eight neighbors likewise have relating pixel esteem. These pixel qualities are then looked into. At that point, the pixels along a circle, i.e. clockwise or counter-clockwise are utilized and we put "1" if the middle pixel's esteem is more prominent than the neighbor; else, we put "0". In this manner, the aftereffect of this an 8-digit paired number is delivered which is then changed over into a decimal esteem. This technique is then rehashed for each pixel in the square. After that in light of its recurrence of every "number" it happens, the histogram is figured over the piece lastly the standardized histograms are connected for every one of the squares. This gives the surface element vector to the information outline

11. Examples of texture primitives

(Source:https://www.hindawi.com/journals/mse/2015/948960/ site)

Then mBm based texture feature extraction is discussed. Texture, characterized as an element of spatial change in pixel powers (dim values), and utilized in various applications has been a field of thorough study by rich scientists. Ojala et al. (1998) have started a fiery surface portrayal method known as the first LBP administrator (Patel & Meshram 2007). In this system, the picture is divided into an arrangement of pieces. For all pixels in a piece, the pixel to each of its 8 neighbors (to its left side top, left-center, left-base, right beat, and so forth.) is differentiated. In this manner, the pixels along a circle, i.e. clockwise or counter-clockwise are sought after and we put "1" if the middle pixel's esteem is more noteworthy than the neighbor; if not, we put "0". This makes accessible an 8-digit double number which is changed into decimal esteem and the business, as usual, is repeating for each pixel in the square. Thusly, the histogram is figured over the square predictable with its recurrence of every "number" happening and to close, standardized histograms of all pieces are linked. This offers the surface properties vector for the info outline.

In this subdivision, clarified a skilled credits extraction figuring to haul out the surface properties by strategy for multi-fractal Brownian development (mBm). Here, mBm is used to look into the sporadic surface assortments of data picture for vivacious surface depend on qualities extraction. The surface depends on qualities can be made by resulting strides as portray underneath: At the start, the every edges or picture are changed into 8×8 squares. Subsequently, the DWT is utilized on every piece to concentrate high and low recurrence data, which offers a first level disintegrated picture of one ballpark picture (LL) and three detail pictures (LH, HL, HH). Here, haar wavelet is utilized to see the high and low recurrence properties from every piece. Haar wavelet change is equipped of recognizing and epitomizes clear marvels in time and recurrence planes. Additionally, it is a smooth and swiftly blurring wavering capacity with a decent restriction in both recurrence and time.

The shape is a basic visual attribute and it is one of the essential credits utilized to depict image content. Then again, shape exhibition and depiction are a simpler said than done task. This is the reason one measurement of protest data is gotten sidetracked when a 3-D certifiable question is anticipated onto a 2-D picture plane. Thusly, the shape removed from the image just in part compares to the anticipated protest. To make the quandary significantly more composite, the shape is over and over besmirched with commotion, abandons self-assertive contortion and impediment. In a general sense, two brands of shape properties which are for the most part utilized can be utilized to symbolize an image for image recovery. One is formeddepend on traits and the extra is area depend on properties. The past is just typified by the edge focuses from a shape, while, the second is include by every single inward point from a shape. Larger scale images need more comparison, which needed more instruction cycle.

MFCC and Pitch are the audio features extracted in this scenario. To extract MFCC, pre-emphasis is done to emphasize high-frequency part by passing through a filter, which will increase the energy of signal at higher frequency. Pre-emphasized signal is divided into frames. In order to keep the continuity of the first and the last points in the frame, each frame is multiplied with a hamming window. To obtain the magnitude frequency response of each frame we perform FFT, which is multiplied by a set of 20 triangular band pass filters in order to get smooth magnitude spectrum. It also reduces the size of features involved. DCT is applied on the output of the triangular band pass filters to get mel-scale cepstral coefficients (Shika

Gupta & Mohd Suhel 2015).To extract pitch, first fast Fourier transform (FFT) is applied on the framed audio signal. By calculating the log spectrum, we have compressed the dynamic range and reduced amplitude differences in the harmonics. We can now return to the time domain through the inverse FT. The peak picking scheme is to determine the cepstral peak in the interval [1 - 20 ms] (Nandhini & Shenbagavalli 2014).

12. Optimization in Video Retrieval

The optimization algorithm is an algorithm used to select the optimal features of the video. In this section, we explain some of the papers used in video retrieval process based on the optimization. In Bae-Muu Chang et al. (2013) have explained a content-based image retrieval method using three kinds of visual features and distance measurements. Particle swarm optimization algorithm is used. For convenience sake, here it was called the CBIRVP method. First, the CBIRVP method extracts three kinds of features: shape, texture, and color features of images. Subsequently, appropriate distance measurements are employed for each kind of features to calculate the similarities between a query image and others in the database. Moreover, PSO algorithm was utilized to optimize the CBIRVP method via searching for nearly best possible combinations of features and their corresponding similarity measurements.

Salahuddin et al. (2012) explained the Content Based Video Retrieval Using Particle Swarm Optimization. This paper provided an efficient methodology that led to incremental improvement in the video search results against a user's query image. Their method employs Particle Swarm Optimization (PSO), an evolutionary population-based search algorithm seeking frames within the video library. The fitness of each swarm particle was the degree of similarity with respect to the video frame(s) fetched through PSO and the content present in the input image provided by the user. This prevents us from the exhaustive linear search of every frame of every video in the library. The relatively best match in each generation of PSO was shown to the user for his engagement.For calculating the fitness of each swarm particle we have tested three similarity measures, 1) convolution 2) correlation based template matching and 3) score from scale-invariant feature transform (SIFT) algorithm. Norlina Mohd Sabri et al. (2013)have explained GSA is a recent algorithm that has been inspired by the Newtonian's low of motion and law of gravity and motion. The algorithm has also been used in many areas. GSA mainly for those researchers who are interested to explore the algorithm's capabilities and performances. Time taken to complete iterations by GSA is better than PSO for the same work.

13. Relevance Feedback for Video Retrieval

Here, the video retrieval process based on relevance feedback is explained. Ionuţ Mironica et al. (2016) have explained the framework for the Fisher Kernel (FK). In this relevance feedback is used. Specifically, they train a Gaussian Mixture Model (GMM) and use this to create an FK representation, which was therefore specialized in modeling the most relevant examples. They use the FK representation to explicitly capture temporal variation. While the GMM was being trained, a user selects from the top examples those which were looking for. This feedback was used to train a Support Vector Machine on the FK representation.Additionally, Yimin Wu & Aidong (2004) have explained the Interactive pattern analysis for relevance feedback in multimedia information retrieval. To perform interactive pattern analysis, they introduced two online pattern classification methods, one is interactive random forests (IRF) and another is adaptive random forests (ARF) that adapt a composite classifier known as random forests for relevance feedback. During interactive multimedia retrieval, both ARF and IRF faster than RRF while achieving considerable precision and recall against the latter.

Similarly, Zimian Li & Ming Zhu (2013) have explained the Light-weight Relevance Feedback Solution for Large-Scale Content-Based Video Retrieval. This paper gives the problem of content-based large scale video retrieval with relevance feedback. They analyzed the common methods to extract feature descriptors from video collections and perform indexing and then retrieval of feature vectors. Instead of knowing similarity-preserving codes, we introduce the relevance feedback approach in an easy way. At descriptor level, a relevance model was explained to merge semantic similarity with the original distance matching. By knowing several weights using canonical correlation analysis (CCA), the videos retrieved changes according to relevance feedback. In Muneesawang & Guan (2003) explained the Automatic relevance feedback for video retrieval. Initially, a representation based on a template frequency model was demonstrated (TFM) that allows full use of the temporal dimension. Then they integrated the TFM with a self-training neural network structure which adaptively captured different degrees of visual importance in a video sequence.

14. Other Approaches Related to Content based Video Retrieval

In this section, we explain some of the papers related to content-based video retrieval system. Mohsen Ramezani & Farzin Yaghmaee (2016) explained the video recommendation system based on efficient retrieval of human actions. The

Recommender Systems (RSs) helped to find the users' most favorite items and finding these items relies on items or users similarities. The users start to impress the recommendation quality by many factors like sparsity and cold. In some systems, attached tags were used for searching items (e.g. videos) as a personalized recommendation. Here, content-based search was used for finding items (here, videos are considered) where a video was taken from the user to find and recommend a list of most similar videos to the query one. Since most videos relate to humans, a low complex scalable method was presented to recommend videos based on the model of included action. In order to model human actions, some interest points were extracted from each action and their motion information used to compute the action representation.Moreover, Rafal Kapela et al. (2015) explained the real-time event classification in field sports videos where they presented based on new global image descriptors how the same underlying audio–visual feature extraction algorithm was robust across a range of different sports. This alleviated the need to tailor it to a particular sport. In addition to this, three different classifiers were introduced and evaluated in order to detect events using these features: an Elman neural network, a feed-forward neural network, and a decision tree. Each was investigated and evaluated in terms of their usefulness for real-time event classification and also a ground truth dataset together with an annotation technique was explained for performance evaluation of each classifier. This was useful to others interested in this problem.

Similarly, Jaesik Choi et al. (2013) explained the spatiotemporal pyramid matching for video retrieval, for a given query video clip where they find the most relevant video segments from video database. Finding relevant video clips was a complex task because objects in a video clip were constantly moved over time. To perform this task efficiently, a video matching called Spatio-Temporal Pyramid Matching (STPM) was explained. Considering features of objects in 2D space and time, STPM divides a video clip into a 3D spatiotemporal pyramidal space recursively. In different resolutions, the features are compared. In order to improve the retrieval performance, both static and dynamic features of objects are considered which provides a sufficient condition in which the matching gets the additional benefit from the temporal information.In Di Zhong & Shih-Fu Chang (2004) have explained the Real-time view recognition and event detection for sports video. In sports video, such high-level structures were often characterized by the specific views (e.g., pitching or serve) and within each temporal structural segment the subsequent temporal transition patterns. They were developed robust statistical models which helped to detect the domain-specific views with real-time performance and high accuracy. The models combine domain-specific constraints on the spatiotemporal properties of the segmented regions (e.g., locations, shapes, and motion of the objects) and domain-independent global color filtering method. The real-time performance was accomplished using object-level processing on filtered candidates alone and efficient compressed-domain processing at the front end. High-level events (e.g., baseline plays, strokes, and net plays) were also detected after the view recognition.

A content-based multi-functional video retrieval (MFVR) system was proposed by Huang-Chia Shih & Chung-Lin Huang (2005). Initially, a hierarchical retrieval technique was discussed and the system feasibility demonstrated via two case-studies. To proceed further, different content semantics based methodology of content analysis was also presented. At last, the efficacy of MFVR in creating diverse summarized video is revealed by experiments. Lili (2009) proposed a framework for video retrieval based on repeated sequence of navigating, searching, browsing, and viewing. High-level and low-level visuals are processed by a framework of structural video analysis. Here, HMM has been employed as the key content-based retrieval processes. The proposed framework also utilizes the video and audio extracted features. Obtaining answers to the questions was made easier and faster by recognizing high-level semantic actions and encoding more semantic details using the semantic features.

Maheshkumar et al. (2010) has offered an approach towards automated highlight generation of broadcast sports video series from its extorted events and semantic ideas. A sports video was hierarchically separated into temporal partitions namely, mega slots, slots, and semantic entities, namely concepts, and events. In every series, categorization is done based on the concept of sequential association mining in the planned technique where event sequence is extorted from the video. The extorted ideas and events within the concepts were selected according to their extent of significance to comprise those in the highlights. A factor level of concept was planned, which gave an option to the user about the swiftness of the extorted ideas created for particular highlight duration. There is effective extortion of highlights from recorded video of a cricket match. Moreover Ork de Rooij & Marcel Worring (2013) have implemented a system for active bucket-based video retrieval, evaluating two different learning strategies. The method shows that it used in video retrieval with an evaluation using three groups of non-expert users. One baseline group is used only for the categorization of features of Media. Huan-Bo Luan et al. (2011) have introduced an effective interactive video retrieval system named Vision Go. It was a joint exploration of the human and computer to accomplish video retrieval with high effectiveness and efficiency.

Similarly, Thanga Ramya & Rangarajan (2011) using different knowledge-based methods addressed the specific aspect of extracting semantics from raw video data. In particular, this paper focuses on three techniques namely, Dynamic Bayesian Networks (DBNs), rules and Hidden Markov Models (HMMs), First, an approach that supports spatio-temporal formalization of high-level concepts was introduced. Then the focus of this paper shifted towards stochastic methods. It also demonstrates how HMMs and DBNs could be effectively used for content-based video retrieval from multimedia databases. Although this work has not compared the two stochastic approaches between each other, an intuitive conclusion was that the DBN approach has been more suitable for fusing multimodalities in retrieval. Similarly, Yanqiang et al. (2012) have explained the Matching of Video Sequence based on the Color Correlation Invariance. Here, they retrieve the system based on the color correlation features.

II. ACKNOWLEDGEMENT

The author thanks Dr.S.Sridhar, Professor and Vice Chancellor, Dr.K.N.Modi University, for communicating this article to this Journal for publication.

III. REFERENCES

1. Abdelati Malek Amel, Ben Abdelali Abdessalem & Mtibaa Abdellatif 2010, ‘Video shot boundary detection using motion activity descriptor’, Journal of Telecommunications, vol. 2, no. 1.
2. Abinaya Sambat Kumar & Nirmala A 2015, ‘A Survey on Multimodal Techniques in Visual Content Based Video Retrieval’, International Journal of Advanced Research in Computer Science and Software Engineering, vol. 5, no. 1.
3. Alberto Del Bimbo & Marco Bertini 2007, ‘Multimedia Ontology Based Computational Framework for Video Annotation and Retrieval’, Springer-Verlag Berlin Heidelberg.
4. Anastasios D Doulamis, Nikolaos D Doulamis & Stefanos D Kollias 2000, ‘A fuzzy video content representation for video summarization and content-based retrieval’, Journal of Signal Processing, vol. 80, no. 6, pp. 1049-1067.
5. Archana V Potnurwar & Mohammad Atique 2014, ‘Visual Attention Key Frame Extraction for Video Annotations’, International Journal of Computer Science Engineering, vol. 3, no.1.
6. Aree A Mohammed 2008, ‘Region-Based Segmentation of Generic Video Scenes Indexing’, Proceedings of World Academy of Science, Engineering and Technology, vol. 32.
7. Arevalillo-Herrez M, Ferri FJ & Domingo J 2010, ‘A naive relevance feedback model for content-based image retrieval using multiple similarity measures, Pattern Recognition, vol. 43, pp. 619-629.
8. Babaguchi N, Ishida T & Morisawa K 2004, ‘Scene retrieval with sign sequence matching based on video and audio features’, IEEE International Conference, vol. 2.
9. Babaguchi N, Kawai Y & Kitahashi T 2002, ‘Event based indexing of broadcasted sports video by intermodal collaboration’, Multimedia, IEEE Transactions, vol. 4, pp. 68-75.
10. Bae-Muu Chang, Hung-Hsu Tsai & Wen-Ling Chou 2013, ‘Using visual features to design a content-based image retrieval method optimized by particle swarm optimization algorithm’, Engineering Applications of Artificial Intelligence, vol. 26, pp. 2372-2382.
11. Bharat K & Henzinger M 1998, ‘Improved algorithms for topic distillation in hyperlinked environments’, Proceedings of ACM-SIGIR’98, Melbourne, pp. 104-111.
12. Bole R, Yeo B & Yeung M 1998, ‘Video query: research directions’, IBM Journal of Research and Development, vol. 42, no. 2, pp. 233-252.
13. Brin S & Page L 1998, ‘The anatomy of a large-scale hyper textual Web search engine’, Proceedings of WWW7, Brisbane, pp. 107-117
14. Changsheng Xu 2008, ‘Using Webcast Text for Semantic Event Detection in Broadcast Sports Video’, IEEE Transactions on Multimedia, vol. 10, no. 7, pp. 1342-1355.
15. Chasanis V, Likas A & Galatsanos N 2007, ‘Scene detection in videos using shot clustering and symbolic sequence segmentation’, in IEEE Workshop on Multimedia Signal Processing.

16. Che-Yen Wen, Liang-Fan Chang & Hung-Hsin Li 2007, 'Content based Video Retrieval with Motion Vectors and the RGB Color Model', Forensic Science Journal, vol. 6, no. 2, pp. 1-36.

17. Chih-Chin Lai & Ying-Chuan Chen 2011, 'A User-Oriented Image Retrieval System Based on Interactive Genetic Algorithm', IEEE Trans on Instrumentation and Measurement, vol. 60, no. 10.

18. Chih-Wen Su, Hong-Yuan Mark Liao & Kuo-Chin Fan 2005, 'A Motion-Flow-Based Fast Video Retrieval System', 7th ACM SIGMM international workshop on Multimedia Information retrieval.

19. Chinmaya Misra & Shamik Sural 2006, 'Content Based Image and Video Retrieval Using Embedded Text', Lecture Notes in Computer Science, vol. 3852, pp. 111-120.

20. Choi KC, Ko YM, Cheon GY, Kim HI, Shin SY & Rhee YW 2006, 'Video shot boundary detection algorithm', Computer. Vis., Graph. Image Process, Lect. Notes Computer Sci., vol. 4338, pp. 388-396.

21. Dagtas S, Al-Khatib W, Ghafoor A & Kashyap R 2000, 'Models for motion-based video indexing and retrieval', IEEE Transactions on Image Processing, vol. 9, no. 1, pp. 88-101.

22. Deepali Bhawarthi & Shriniwas Gadage 2012, 'Enriching Feature Extraction for Cricket Video Event Detection', International Journal of Engineering Research and Technology (IJERT), vol. 1, no. 7.

23. Di Zhong & Shih-Fu Chang 2004, 'Real-time view recognition and event detection for sports video', Journal of Visual Communication and Image Representation, vol. 15, no. 3, pp. 330-347.

24. Dimitrova Rui, Sethi K & Rahman SM (ed) Idea group publishing.

25. Drelie Gelasca E & Ebrahimi T 2009, 'On Evaluating Video Object Segmentation Quality: A Perceptually Driven Objective Metric', IEEE Journal of Selected Topics in Signal Processing, vol. 3, no. 2, pp. 319-335.

26. Dubravko Culibrk, Oge Marques, Daniel Socek, Hari Kalva & Borko Furht 2007, 'Neural Network Approach to Background Modeling for Video Object Segmentation', IEEE Transactions on Neural Networks, pp. 1614-1627.

27. Dyana A & Das S 2010, 'MST-CSS (Multi-Spectrum-Temporal Curvature Scale Space), a Novel Spatio-Temporal Representation for Content-Based Video Retrieval', IEEE Transactions on Circuits and Systems for Video Technology, vol. 20, no. 8, pp. 1080-1094.

28. Ekin A & Tekalp M 2003, 'Automatic Soccer Video Analysis and Summarization', EEE Transaction on Image Processing, vol. 12, pp. 796- 807.

29. Ganesh I Rathod & Dipali A Nikam 2014, 'Review on Event Retrieval in Soccer Video', International Journal of Computer Science and Information Technologies, vol. 5, no. 4, pp. 5601-5605.

30. Geetha P & Vasumathi Narayanan 2008, 'A Survey of Content-Based Video Retrieval', Journal of Computer Science, vol. 4, no. 6, pp. 474-486.

31. Gijesh Varghese & Zhou Wang 2010, Video Denoising Based on a Spatiotemporal Gaussian Scale Mixture Model', IEEE transactions on circuits and systems for video technology, vol. 20, no. 7.

32. Gordon M & Pathak P 1999, 'Finding information on the World Wide Web: the retrieval effectiveness of search engines', Information Processing & Management, vol. 35, no. 2, pp. 141-180.

33. Guozhu Liu & Junming Zhao 2009, 'Key Frame Extraction from MPEG Video Stream', Proceedings of the Second Symposium International Computer Science and Computational Technology, pp. 007-011.

34. Hanjalic A 2002, 'Shot-boundary detection: unraveled and resolved', IEEE Transactions on Circuits and Systems for Video Technology, vol. 12, no. 2, pp. 90-105.

35. Haojin Yang & Meinel C 2014, 'Content Based Lecture Video Retrieval Using Speech and Video Text Information', IEEE Transactions on Learning Technologies, vol. 7, no. 2, pp. 142-154

36. Heng WJ & Ngan 2002, 'Shot boundary refinement for long transition in digital video sequence', Multimedia, IEEE Transactions, vol. 4, pp. 434-445.

37. Huang-Chia Shih & Chung-Lin Huang 2005, 'Content-based multi-functional video retrieval system', International Conference on Consumer Electronics.

38. Ionuţ Mironica, Bogdan Ionescu, Jasper Uijlings & Nicu Sebe 2016, ‘Fisher Kernel Temporal Variation-based Relevance Feedback for video retrieval’, Journal of Computer Vision and Image Understanding, vol. 143, pp. 38-51.

39. Jaesik Choi, Ziyu Wang, Sang-Chul Lee & Won J. Jeon 2013, ‘A spatio-temporal pyramid matching for video retrieval’, Journal of Computer Vision and Image Understanding, vol. 117, no. 6, pp. 660-669.

40. Ja-Hwung Su, Yu-Ting Huang, Hsin-Ho Yeh & Vincent S Tseng 2010, ‘Effective content-based video retrieval using pattern-indexing and matching techniques’, Journal of Expert Systems with Applications, vol. 37, no. 7, pp. 5068-5085.

41. Jordi Pont-Tuset, Miquel A Farre & Aljoscha Smolic 2015, ‘Semi-Automatic Video Object Segmentation by Advanced Manipulation of Segmentation Hierarchies’, International Workshop on Content-Based Multimedia Indexing.

42. Jun Wei Hsieh, Shang-Li Yu & Yung-Sheng Chen 2006, ‘Motion-based Video Retrieval by Trajectory Matching’, IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, no. 3, pp. 396-409.

43. Khin Thandar Tint & Kyi Soe 2013, ‘Key Frame Extraction for Video Summarization Using DWT Wavelet Statistics’, International Journal of Advanced Research in Computer Engineering and Technology (IJARCET) vol. 2, no 5.

44. Kimiaki Shirahama, Kazuyuki Otaka & Kuniaki Uehara 2007, ‘Content-Based Video Retrieval Using Video Ontology’, in proceedings with Ninth IEEE International Symposium on Multimedia Workshops (ISMW 2007), pp. 283-289.

45. Kin-Wai Sze, Kin-Man Lam & Guoping Qiu 2005, ‘A New Key Frame Representation for Video Segment Retrieval’, IEEE Transactions on Circuits and Systems for Video Technology, vol. 15, no. 9.

46. Kleinberg J 1998, ‘Authoritative sources in a hyperlinked environment’, Proceedings of 9th ACM-SIAM Symposium on Discrete Algorithms, pp. 668-677

47. Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik & Karen Egiazarian 2007, ‘Image demonising by sparse 3D transform-domain collaborative filtering’, IEEE Transactions on Image processing, vol. 16, no. 8.

48. Lakshmi Priya GG & Domnic S 2014, ‘Shot based key frame extraction for ecological video indexing and retrieval’, Journal of Ecological Informatics, vol. 23, pp. 107-117.

49. Lawrence S & Lee Giles C 1999, ‘Accessibility of information on the Web’, Nature, vol. 400, no. 6740, pp. 107-110.

50. Lei Xu & Kongqiao Wang 2008, ‘Extracting text Information for Content-based Video Retrieval’, In Proceedings of the 14th International conference on Advances in multimedia modeling, pp. 58-69.

51. Leighton HV & Srivastava J 1999, ‘First 20 precision among World Wide Web search services (search engines)’, Journal of the American Society for Information Science, vol. 50, no. 1, pp. 870-881.

52. Lew M, Sebe N & Gardner P 2001, ‘Video indexing and understanding’, M. Lew (Ed.), Principles of Visual Information Retrieval, Springer, Berlin, pp. 163-196.

53. Li B & Ibrahim Sezan M 2001, ‘Event detection and summarization in sports video’, Presented at Content-Based Access of Image and Video Libraries, IEEE Workshop on, Sharp Labs of America, Camas, WA, USA.

54. Lili NA 2009, ‘Hidden Markov Model for Content-Based Video Retrieval’, In proceedings ofThird Asia International Conference (AMS '09), vol. 25, pp. 353-358.

55. Liu GH, Li ZY, Zhang L & Xu Y 2011, Image retrieval based on microstructure descriptor, Pattern Recognition, vol. 44, pp. 2123-2133.

56. Liu Huayong 2004, ‘Content-Based TV Sports Video Retrieval Based on Audio-Visual Features and Text Information’, IEEE/WIC/ACM International Conference.

57. Liya Thomas & Syama 2014, ‘Ontology Based Video Annotation and Retrieval System’, International Journal of Emerging Technology and Advanced Engineering, vol. 4, no. 7.

58. Lu ZM & Shi Y 2013, ‘Fast Video Shot Boundary Detection Based on SVD and Pattern Matching’, IEEE Transactions on Image Processing, vol. 22, no. 12.

59. Luan H, Zheng Y, Wang M & Chua 2011, ‘Vision Go: Towards video retrieval with joint exploration of human and computer', Journal of Information Science, vol. 181, no. 19, pp. 4197-4213.

60. Maheshkumar H Kolekar & Somnath Sengupta 2010, 'Semantic concept mining in cricket videos for automated highlight generation', Multimedia Tools Appl, vol. 47, pp. 545-579.

61. Marchiori M 1997, 'The quest for correct information on the Web: Hyper search engines', Proceedings of WWW6, Santa Clara (CA).

62. Ming Zhao, Jiajun Bu & Chun Chen 2002, 'Semi-Automatic Video Object Segmentation Basing On Hierarchy Optical Flow', in SPIE: Electronic Imaging and Multimedia Technology III, vol. 4925, pp. 307-316.

63. Mohsen Ramezani & Farzin Yaghmaee 2016, 'A novel video recommendation system based on efficient retrieval of human actions', journal of Statistical Mechanics and its Applications.

64. Muneesawang P & Guan L 2003, 'Automatic relevance feedback for video retrieval', in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., vol. 3.

65. Nagarajan G & Minu RI 2015, 'Fuzzy Ontology Based Multi-Modal Semantic Information Retrieval', International Conference on Computer, Communication and Convergence, vol. 48, pp. 101-106.

66. Nandhini S & Shenbagavalli A 2014, 'Voiced/Unvoiced Detection using Short Term Processing', International Journal of Computer Applications (0975-8887).

67. Nepal S, Srinivasan U & Reynolds G 2001, 'Automatic detection of 'Goal' segments in basketball videos', presented at ACM International Conference on Multimedia, Ottawa, Canada.

68. Nicu Sebe, Michael S Lew & Arnold W M Smeulders 2003, 'Video retrieval and summarization: editorial introduction', Computer Vision and Image Understanding, vol. 92, no. 2-3, pp. 141-146.

69. Nishan Singh & Vijay Laxmi 2014, 'Audio Noise Reduction from Audio Signals and Speech Signals', International Journal of Computer Science Trends and Technology, vol. 2, no. 5.

70. **Padmakala S** & Anandha Mala GS2010, 'Novel Video Object Segmentation Approach for Noisy Video Sequences towards Effective Video Retrieval', International Journal of Computer Theory and Engineering, vol. 2, no. 6, pp. 1793-8201.

71. **Padmakala S**, Anandha Mala GS& Shalini M 2011, 'An effective content based video retrieval utilizing texture, color and optimal key frame features', International Conference on Image Information Processing (ICIIP).

72. **Padmakala S** & Anandha Mala GS2012, 'A Technique to Content-Based Video Retrieval Utilizing Diverse Set of Features', European Journal of Scientific Research, ISSN 1450-216X, vol. 83, no. 4, pp. 558-575 .

73. **Padmakala S**,Muthuchelvi P & Anandha Mala GS 2014, 'IVRSC: An Interactive and Intelligent Video Retrieval System for Cricket Videos Using Multi-Features', International Journal of Applied Engineering Research, ISSN 0973-4562, vol. 9, no. 24, pp. 27457-27491.

74. **Padmakala S** & Anandha Mala GS2018, 'Interactive Video Retrieval Using Semantic Level Features and relevant feedback', International Arab Journal of Information Technology,ISSN 2309-4524, vol. 13, no. 1.

75. Patel BV & Meshram BB 2007, 'Retrieving and Summarizing Images from PDF Documents', International Conference on Soft computing and Intelligent Systems.

76. Peng Tang & Lin Gao 2008, 'Video Object Segmentation Based On Graph Cut With Dynamic Shape Prior Constraint', 19th International Conference on Pattern Recognition (ICPR), pp. 1-4.

77. Ponceleon D, Srinivasan S, Amir A Petkovic & Diklic 1998, 'Key to effective video retrieval: Effective cataloguing and browsing', in IEEE International Workshop on Content-based image and video databases, IEEE Computer Society, pp. 99-107.

78. Rafal Kapela, Aleksandra Świetlicka, Andrzej Rybarczyk, Krzysztof Kolanowski & Noel E.Connor 2015, 'Real-time event classification in field sport videos', Journal of Signal Processing: Image Communication, vol. 35, pp. 35-45.

79. Ramya S & Rangarajan 2011, 'Knowledge Based methods for Video Data Retrieval', International Journal of Computer Science and Information Technology, vol. 3, no. 5, pp. 469-481

80. Ren W, Singh S, Singh M & Zhu YS 2009, State-of-the-art on spatiotemporal information-based video retrieval, Pattern Recognition, vol. 42, pp. 267-282.

81. Rooij O & Worring 2013, 'Active Bucket Categorization for High Recall Video Retrieval', IEEE Transactions on Multimedia, vol. 15, no. 4.

82. Rui Y, Gupta A & Acero A 2000, 'Automatically extracting highlights for TV Baseball programs', presented at ACM International Conference on Multimedia, Marina Del Rey, California, and United States.

83. Ryen W White, Ian Ruthven & Joemon M Jose 2005, 'A Study of Factors Affecting the Utility of Implicit Relevance Feedback'.

84. Rynson WH Lau, Qing Li & Antonio Si 2000, 'Video MAP: A Generic Framework for Video Management and Application processing', 33rd Hawaii International Conference on System Sciences (HICSS), pp. 8044, vol. 8.

85. Sadlier DA & O'Connor Nebraska 2005, 'Event detection in field sports video using audio-visual features and a support vector Machine', IEEE Transactions on Circuits and Systems for Video Technology, vol. 15, no. 10.

86. Salahuddin A, Naqvi A, Murtaza K & Akhtar J 2012, 'Content Based Video Retrieval Using Particle Swarm Optimization', Frontiers of Information Technology (FIT), 2012 10th International Conference, pp. 79-83.

87. Salton G 1989, 'Automatic text processing, the transformation, analysis, and retrieval of information by computer', Addison-Wesley, Reading (MA).

88. Sankar KP, Pandey S & Jawahar CV 2006, 'Text Driven Temporal Segmentation of Cricket Videos', in Proceedings Sixth of Indian Conference on Computer Vision, Graphics and Image Processing, Madurai, India, pp. 433-444.

89. Schwartz C 1998, 'Web search engines', Journal of the American Society for Information Science, vol. 49, no. 11, pp. 973-982.

90. Sebastien Lefevrea, Jerome Hollera & Nicole Vincenta 2003, 'A review of real-time segmentation of uncompressed video sequences for content-based search and retrieval', Real-journal of Time Imaging, vol. 9, no. 1, pp. 73-98.

91. Seung-Hoon Han, Kuk-Jin Yoon & In So Kweon 2000, 'A new technique for shot detection and key frames selection in histogram space' In: 12th Workshop on Image Processing and Image Understanding, pp. 475-479.

92. Shahraray B 1999, 'Multimedia information retrieval using pictorial transcripts', B. Furth (Ed.), Handbook of Multimedia Computing, CRC Press, Boca Raton, FL, pp. 345-359.

93. Shanmugham TN & Priya Rajendran 2009, 'An Enhanced Content-Based Video Retrieval System Based on Query Clip', International Journal of Research and Reviews in Applied Sciences, vol. 1, no. 3.

94. Shika Gupta & Mohd Suhel 2015, 'Speech Recognition using MFCC & VQ', International Journal of Scientific Engineering And Technology Research, vol. 04, no. 01, pp. 0058-0061.

95. Shweta Ghodeswar & Meshram BB 2010, 'Content Based Video Retrieval', Proceeding of International Symposium on Computer Engineering and technology, Mandi Gobindgarh, Punjab.

96. Simon Jones & Ling Shao 2013, 'Content-based retrieval of human actions from realistic video databases', Journal of Information Sciences, vol. 236, 1, pp. 56-65.

97. Song X & Fan G 2006, 'Joint Key-Frame Extraction and Object Segmentation for Content-Based Video Analysis', IEEE Transactions on Circuits and Systems for Video Technology vol. 16, no. 7.

98. Su CW, Liao HYM, Tyan HR, Lin CW, Chen DY & Fan KC 2007, 'Motion Flow-Based Video Retrieval', journal of IEEE Transactions on Multimedia, vol. 9, no. 6, pp. 1193-1201.

99. Sudhir S Kanade & Pradeep M Patil 2012, 'Lawn Tennis Video Summarization based on Audiovisual and Text Feature Analysis', International Journal of Computer Applications, vol. 42, no. 19.

100. Sun X, Zhao L & Zhang M . 2011, 'A Novel Shot Boundary Detection Method Based on Genetic Algorithm-Support Vector Machine', Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2011 International Conference, vol. 1, pp. 144-147.

101. Teraguchi M, Masumitsu K, Echigo T, Sekiguchi S & Etoh M 2002, 'Rapid generation of event-based indexes for personalized video digests', Presented at Pattern Recognition.

102. The open video project from 'http://www.open-video.org/.

103. Thompson R 1998, 'Grammar of the Shot', Focal Press.

104. Tiburzi F Escudero, Bescos M & Martinez 2008, 'A Ground Truth for Motion-Based Video-Object Segmentation', 15th IEEE International Conference on Images Processing (ICIP), pp. 17-20.

105. Torres RDS, Falcao AX, Goncalves MA, Papa JAP, Zhang B, Fan W & Fox EA 2009, 'A genetic programming framework for content-based image retrieval', Pattern Recognition, vol. 42, pp. 283-292.

106. Vijayakumar V & Nedunchezhian R 2012, 'Event detection in cricket video based on visual and acoustic features', Journal of Global Research in Computer Science, vol. 3, no. 8.

107. Walid G Aref, Ann C Catlin, Ahmed K Elmagarmid J Fan, Moustafa A Hammad, Ihab Ilyas, Mirette Marzouk & Thanaa Ghanem 2003, 'Video Query Processing in the VDBMS Test bed for Video Database Research', 1st ACM International Workshop on Multimedia Databases, (ACM MMDB), pp. 25-32.

108. Wattanarachothai W & Patanukhom K 2015, 'Key frame extraction for text based video retrieval using Maximally Stable Extremal Regions', Industrial Networks and Intelligent Systems, pp. 29-37.

109. Wu C, Ma YF, Zhang HJ & Zhong YZ 2002, 'Events recognition by semantic inference for sports video', Presented at Multimedia and Expo.

110. Xiaoming Liu & Tsuhan Chen 2004, 'Shot Boundary detection using temporal statistics modeling', Acoustics, Speech, and Signal Processing IEEE International.

111. Xu Zhang, Bao-Long Guo,Guiyue Zhang & Yunyi Yan 2011, 'An Image Retrieval Method Based on r/KPSO', IEEE Second International Conference on Innovations in Bio-inspired Computing and Applications, pp. 69-72, DOI 10.1109/IBICA.2011.22.

112. Xuemei Wang, Dengyin Zhang, Min Zhu, Yingtian Ji & Jin Wang2015', Improved Image Denoising Based on 3D Collaborative Filtering', International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 8, no. 4, pp. 227-236.

113. Yanqiang Lei, Weiqi Luo, Yuangen Wang & Jiwu Huang 2012, 'Video Sequence Matching Based on the Invariance of Color Correlation', IEEE transactions on circuits and systems for video technology, vol. 22, no. 9.

114. Yimin Wu & Aidong Zhang 2004, 'Interactive pattern analysis for relevance feedback in multimedia information retrieval', Springer-Verlag, Multimedia Systems, vol. 10, pp. 41-55.

115. Yipei Wang, Shourabh Rawat & Florian Metze 2014, 'Exploring audio semantic concepts for event-based video retrieval', IEEE International Conference on Acoustics, Speech and Signal Processing.

116. Yu Meng, Li-Gong Wang & Li-Zeng Mao 2009, 'A shot boundary detection algorithm based on Particle Swarm Optimization Classifier', Machine Learning and Cybernetics, 2009 International Conference, vol. 3, pp. 1671-1676.

117. Zabih R, Miller J & Mai K 1995, 'A feature based algorithm for detecting and classifying scene breaks', In Proceedings of the 3rd ACM International Conference on Multimedia, pp. 189-200.

118. Zeinik-Manor L & Irani M 2001, 'Event-based analysis of video', Presented at Computer Vision and Pattern Recognition, Proceedings of the 2001 IEEE Computer Society Conference on, The Weizmann Institute of Science.

119. Zhang HJ 1999, 'Content-based video browsing and retrieval', 'CRC press LLC.

120. Zhe Ming Lu & Yong Shi 2013, 'Fast Video Shot Boundary Detection Based on SVD and Pattern Matching', Image processing IEEE Transactions, vol. 22, no. 12.

121. Zhonghua Sun, Jia K & Chen H 2008, 'Video Key Frame Extraction Based on Spatial-Temporal Color Distribution', Intelligent Information Hiding and Multimedia Signal Processing, pp. 196-199.

122. Zimian Li & Ming Zhu 2013, 'A Light-weight Relevance Feedback Solution for Large Scale Content-Based Video Retrieval', IJCSI International Journal of Computer Science Issues, vol. 10, no. 1, pp. 3.

